

Efficient Generation ϵ -close to $G(n, p)$ and Generalizations

Antonio Blanca*

Milena Mihail†

Abstract

We give an efficient algorithm to generate a graph from a distribution ϵ -close to $G(n, p)$, in the sense of total variation distance. In particular, if p is represented with $O(\log n)$ -bit accuracy, then, with high probability, the running time is linear in the expected number of edges of the output graph (up to poly-logarithmic factors). All our running times include the complexity of the arithmetic involved in the corresponding algorithms. Previous standard methods for exact $G(n, p)$ sampling (see e.g. [2]) achieve similar running times, however, under the assumption that performing real number arithmetic with arbitrary accuracy takes constant time. We note that the actual accuracy required by these methods is $O(n)$ -bit per step, which results in quadratic running times. We also note that compromising on the $O(n)$ -bit accuracy requirement causes arbitrary large biases in the sampling.

The main idea of our $G(n, p)$ generation algorithm is a Metropolis Markov chain to sample ϵ -close from the binomial distribution. This is a new method for sampling from the binomial distribution: it is of separate interest and may find other useful applications. Our analysis accounts for all necessary bit-accuracy and arithmetic. Our running times are comparable to known methods for exact binomial sampling (e.g. surveyed in [15]), however, only when the latter do not account for bit-accuracy and assume arbitrary bit arithmetic in constant time. Dropping these assumptions affects their running times and/or causes bias which has never been quantified. In this sense, our work can be viewed as a rigorous quantification of the tradeoff between accuracy and running time, when all computational aspects are taken into account.

We further obtain efficient generation algorithms for random graphs with given arbitrary degree distributions, Inhomogeneous Random Graphs when the kernel function is the inner product, and Stochastic Kronecker Graphs. Efficient generation of these random graph models is essential for modeling large scale complex networks. To the best of our knowledge, our work can be viewed as the first effort to simulate efficient generation of graphs from classical random graph models, while taking into account implementational considerations as fundamental computational aspects, and quantifying the tradeoff between accuracy and running time in a way that can be useful in practice.

*U.C. Berkeley, Email:ablanca@eecs.berkeley.edu, Supported in part by NSF-CCF-TF-0830683.

†Georgia Tech, Email:mihail@cc.gatech.edu Supported in part by NSF-CCF-TF-0830683.

1 Introduction

Let $n \geq 1$ be an integer and let \mathcal{G}_n be the set of $2^{\binom{n}{2}}$ undirected simple graphs on n vertices. For $0 \leq p \leq 1$, we typically think of $G(n, p)$ as a graph in \mathcal{G}_n , where each edge $\{u, v\}$ is present with probability p and absent with probability $(1 - p)$, independently from all other edges. The straightforward way to generate such a graph involves $\binom{n}{2}$ independent experiments. Thus the running time is $\Omega(n^2)$. In practice, we need an additional $\Omega(\log n)$ to represent n distinct vertices and an additional $\Omega(\log p^{-1})$ to simulate sampling with probability p resulting in $\Omega(n^2(\log n + \log p^{-1}))$ total running time.

If m is the number of edges of the output graph, it is highly desirable to aim for $O(m)$ running times, especially when $E[m] \ll n^2$ and n is very large. This is true, for example, in the case of complex networks. In these cases, the number of vertices n is known to scale massively, while the corresponding graphs remain relatively sparse [1, 5, 7, 12, 13] (for example, due to natural underlying resource considerations, as has been discussed extensively in the literature). Taking into account the implementational issues mentioned in the previous paragraph and the resource considerations, we should aim for $O(m(\log n + \log p^{-1}))$ running times.

Let w_1, \dots, w_n be non-negative weights on n vertices. Thus we may refer to the w_u 's as an n -dimensional vector \vec{w} . Let $G(n, \vec{w})$ be a graph in \mathcal{G}_n , where each edge $\{u, v\}$ is present with probability $\min\{w_u w_v, 1\}$, independently from all other edges. Notice that $G(n, p)$ is a special case of $G(n, \vec{w})$, where $w_u = \sqrt{p}$ for all u . Moreover, $G(n, \vec{w})$ subsumes the case of random graphs with given expected degrees [8, 9, 10, 11, 21, 22].

Let $d \geq 1$ be an integer and let $\vec{w}_1, \dots, \vec{w}_n$ be vectors in d -dimensional real space with non-negative coordinates: $w_{uk} \geq 0, \forall 1 \leq u \leq n, \forall 1 \leq k \leq d$. We may thus refer to the \vec{w}_u 's using an $n \times d$ matrix W . Let $G(n, W)$ be a graph in \mathcal{G}_n , where each edge $\{u, v\}$ is present with probability $\min\{\langle w_u, w_v \rangle, 1\}$, independently from all other edges, and where $\langle w_u, w_v \rangle$ is the usual inner product $\langle w_u, w_v \rangle = \sum_{k=1}^d w_{uk} w_{vk}$. Notice that $G(n, \vec{w})$ is a special case of $G(n, W)$, where $d = 1$. Moreover, $G(n, W)$ is a special case of Inhomogeneous Random Graphs [5, 6], where the kernel function is the inner product. Random inner product graphs have been also studied in [26, 27]. In the context of complex networks, the interpretation of vertices represented as d -dimensional vectors is natural. Real datasets are categorical. Therefore, each dimension represents a distinct attribute, and data points connect with probabilities related to their similarity according to such attributes.

Let $k, d \geq 1$ be integers, and let \mathcal{P} be a $d \times d$ initiator matrix. Define recursively the matrix $K_k = \mathcal{P} \otimes K_{k-1}$, where $K_0 = I$ and \otimes is the Kronecker product of matrices [17]. For $n = d^k$, let $G(n, \mathcal{P})$ be a graph in \mathcal{G}_n , where each edge $\{u, v\}$ is present with probability $\min\{K_k(u, v), 1\}$, independently from all other edges. $G(n, \mathcal{P})$ is known in the literature as Stochastic Kronecker Graphs [14, 16, 17, 18, 24].

In the context of complex networks, the model of random graphs with given expected degrees $G(n, \vec{w})$, the model of Inhomogeneous Random Graphs where the kernel function is the inner product $G(n, W)$, and the Stochastic Kronecker model $G(n, \mathcal{P})$ have been used to produce synthetic graphs that capture important structural properties of real networks. Thus, efficient algorithms to generate such random graphs are important both in theory and in practice.

Let π be a probability distribution on \mathcal{G}_n . Random graph models, such as $G(n, p)$, $G(n, \vec{w})$, $G(n, W)$, and $G(n, \mathcal{P})$ defined above, are equivalent to such distributions π over \mathcal{G}_n . Moreover, properties of random graphs in such models are typically expressed as holding with high probability. This means that, for some constant $c > 0$, the subset \mathcal{G}_{BAD} of graphs in \mathcal{G}_n for which the property does not hold has $\pi(\mathcal{G}_{\text{BAD}}) \leq n^{-c}$ [4, 12]. Such quantification is of fundamental predictive value,

both in theory and applications.

Let π' be a probability distribution on \mathcal{G}_n that is ϵ -close to π , in the usual sense of variation distance: $\max_{\mathcal{H} \subset \mathcal{G}_n} |\pi'(\mathcal{H}) - \pi(\mathcal{H})| \leq \epsilon$. In the context of large scale complex networks it is typical to check experimentally for various desired (or not desired) properties of graphs generated according to a target distribution π . If generation was done according to π' that is provably ϵ -close to π , and if the estimate was $\pi'(\mathcal{G}_{\text{BAD}}) \leq \epsilon_0$, we may readily infer that $\pi(\mathcal{G}_{\text{BAD}}) \leq \epsilon_0 + \epsilon$. On the other hand, if the estimate was $\pi'(\mathcal{G}_{\text{BAD}}) > \epsilon_0$, we may readily infer that $\pi(\mathcal{G}_{\text{BAD}}) > \epsilon_0 - \epsilon$. Thus, if we are able to fine-tune ϵ , generation from π' becomes also of fundamental predictive value.

Our algorithms generate graphs from distributions ϵ -close to those implied by $G(n, p)$, $G(n, \vec{w})$, $G(n, W)$, and $G(n, \mathcal{P})$ respectively, at the cost of a multiplicative factor $O(\log \epsilon^{-1})$ in the running time. This implies that we can set $\epsilon = n^{-c}$, for any constant c , thus matching typical high probability statements for random graphs [4, 12]. For historical reasons, we also mention that ϵ -close sampling has been extensively used in theoretical computer science, for example in the context of approximate counting via Monte Carlo Markov chain simulation, among others.

In particular, for $G(n, p)$, the running time is $O(\mu \log n \log \epsilon^{-1} (\log n + \log p^{-1}))$ in expectation where μ is the expected number of edges of the output graph and the factor $(\log n + \log p^{-1})$ accounts for representation and arithmetic. We may obtain a high probability upper bound on the running time at a cost of an additional $O(\log n)$ multiplicative factor.

Finally, we should note that, for the sake of clarity in presentation, we have not tried to optimize the poly-logarithmic factors at all points (especially in Section 5). There are also points where the pseudocode might hint to “difficult” arithmetic (such as computing square roots). In such cases, we comment right below those pseudocodes, as to how these points can be bypassed.

The rest of the paper is organized as follows. In Section 1.1 we quantify the claim that known algorithms for generating Erdos-Renyi $G(n, p)$ graphs with m output edges run in $O(m)$ time. We note that these methods actually require $O(n)$ bits of arithmetic accuracy per step, which clearly results in $O(mn)$ running times, when all computational aspects are considered. Hence these methods for $G(n, p)$ (and hence all generalizations of $G(n, p)$, e.g. see [22]) are not efficient. In Sections 2, 3, 4, and 5 we develop efficient algorithms to generate random graphs ϵ -close to $G(n, p)$, $G(n, \vec{w})$, $G(n, W)$, and $G(n, \mathcal{P})$ respectively.

1.1 Previous work and Contributions

The standard reference for generating a graph from $G(n, p)$ with running time $O(m)$, where m is the number of output edges, is [2]. This algorithm was recently extended to $G(n, \vec{w})$ in [22]. For $G(n, p)$, the idea in [2] is to order vertices and have each vertex decide its “distance” or “jump” to its next neighbor according to the ordering, where “jumping” a vertex corresponds to an edge that is not present in the final output graph. All vertices bypassed by such jumps will not be processed in the corresponding iteration. Thus, the intuition is that the number of actual steps of the algorithm is proportional to the number of edges present in the output graph, suggesting a $O(m)$ running time.

However, we argue below that the implementation of these “jumps”, which involve the simulation of a negative binomial with parameters n and p , immediately introduce the necessity of n -bit accuracy per “jump”. Therefore, the running time becomes $O(nm)$. The authors bypassed this problem by assuming constant running time for real number arithmetic and representation with arbitrary accuracy. If n -bit accuracy is compromised for any poly $\log n$ -bit accuracy, then the bias becomes immediately arbitrary.

The algorithm in [2] works as follows. First, the vertices are ordered, and the main observation is

that, at any given time during an iteration, the probability of generating the next edge after exactly k trials is $(1-p)^{k-1}p$; i.e. waiting times for the edges are geometrically distributed. Let $q = 1-p$; to sample waiting times, each positive integer k is assigned an interval $I_k \subseteq [0, 1)$ of length $q^{k-1}p$. Realize that $\sum_{k=1}^{\infty} q^{k-1}p = 1$, and if the intervals are contiguous starting at 0, then interval I_k ends at $\sum_{i=1}^k q^{i-1}p = 1 - q^k$. Therefore, the waiting times can be sampled by randomly choosing $r \in [0, 1)$ and selecting the smallest k for which $r < 1 - q^k$. The pseudocode from [2] is included below.

```

 $E \leftarrow \emptyset; v \leftarrow 1; w \leftarrow -1;$ 
while  $v < n$  do
    Draw  $r \in [0, 1)$ , uniformly at random;
     $w \leftarrow w + 1 + \lfloor \frac{\log(1-r)}{\log(1-p)} \rfloor;$ 
    while  $w \geq v$  and  $v \leq n$  do
         $w \leftarrow w - v; v \leftarrow v + 1;$ 
    if  $v < n$  then  $E \leftarrow E \cup \{u, w\};$ 
return  $(E);$ 

```

It should be clear now that we need r sampled with accuracy $O(n)$ bits in order to simulate fair sampling from all intervals. If the arithmetic used has accuracy $\alpha(n)$, then all intervals I_k for $k > \alpha(n)$ will not be represented in the sampling. Thus, $\alpha(n) = o(n)$ introduces arbitrary (unquantified) sampling bias on every step of the algorithm. We also note that the arithmetic in [2] involves the computation of discrete logarithms, and this issue has been raised in [25], who however do not offer any solution with quantified performance. In [22], the same issues carried over for $G(n, \vec{w})$.

Furthermore, the authors in [2] list an array of widely used software for random graph generation that implement inefficient algorithms. They note that such software provides running times tolerable for graphs up to tens of thousands of nodes. We remark that current technology requires synthetic data involving much larger number of nodes (e.g. to simulate social networks.)

Of course, the most natural way to sample $G(n, p)$ is to sample the number of edges m from the binomial distribution $B(\binom{n}{2}, p)$, and then choose exactly m out of the $\binom{n}{2}$ edges at random. The main idea of our $G(n, p)$ generation algorithm is a Metropolis Markov chain to sample ϵ -close from the binomial distribution. Our analysis accounts for all necessary bit-accuracy and arithmetic. Our running times are comparable to known methods for exact binomial sampling (e.g., surveyed in [15]), however, only when the latter do not account for bit-accuracy and assume arbitrary bit arithmetic in constant time. Dropping these assumptions affects their running times and/or causes bias which has never been quantified. Our method to sample from the binomial distribution offers a rigorous quantification of the tradeoff between accuracy and running time, when all computational aspects are taken in to account. Therefore, it is of separate interest and may find other useful applications.

Our algorithms for $G(n, \vec{w})$, $G(n, W)$, and $G(n, \mathcal{P})$ offer also efficient running times while taking into account all computational aspects. To the best of our knowledge, our work can be viewed as the first effort to simulate efficient generation of graphs from classical random graph models, while (a) taking into account fundamental implementational considerations and (b) quantifying possible “errors” in a way that can be useful in practice, i.e. with quantified predictive value.

2 Efficient Generation ϵ -close to $G(n, p)$

In this section π is the probability distribution over \mathcal{G}_n implied by $G(n, p)$. In particular, for any specific graph $G(V, E) \in \mathcal{G}_n$, $\pi(G(V, E)) = p^{|E|}(1-p)^{N-|E|}$, where $N = \binom{n}{2}$. It is obvious that $\pi(|E| =$

$k) = \binom{N}{k} p^k (1-p)^{N-k}$, since there are $\binom{N}{k}$ distinct graphs in \mathcal{G}_n with k edges and π assigns to all these graphs the same probability. Thus, if $B(N, p)$ is the binomial distribution with parameters N and p , then $\pi(|E|=k) = B(k; N, p)$.

Therefore, a natural two-step approach to generate a graph $G(V, E)$ according to π is to sample $|E|$ from $B(N, p)$ and generate a random combination of $|E|$ out of N possible edges. The second step can be implemented in time $O(\max\{|E|, n\} \log n)$ (including representation and arithmetic) using, for example, the classic algorithm in [3]. Sampling from $B(N, p)$, however, is much more involved, and it is analyzed in detail in Subsection 2.1.

In Subsection 2.2 we give algorithm **Sample-G**(n, p, ϵ) which uses the **Coupling Markov Chain** (4) to generate $|E|$, and include all implementational aspects (beyond mixing time) that result in efficient sampling from a distribution ϵ -close to $G(n, p)$.

2.1 Markov Chains ϵ -close to Binomial Distributions

Let $B(N, p)$ be a binomial distribution with parameters $N \in \mathbb{N}$ and $0 < p < 1$. There are several methods to sample from the binomial distribution (for a detailed survey see, e.g. [15]). However, we desire a rigorous quantification of the tradeoff between accuracy and running time. Current known techniques do not have this feature.

We design a Markov chain approach to sample from a distribution that is ϵ -close to $B(N, p)$, in the sense of variation distance. Let $\mu = Np$; the Markov chains are Metropolis-Hastings random walks on a segment of the line around $\mu \pm O(\sqrt{\mu \ln \epsilon^{-1}})$, with expected coupling times (convergence rates) $O(\mu \ln \epsilon^{-1})$. The Markov chain is defined in (3) and the coupling which bounds convergence rate is defined in (4). Lemma 4 and Theorem 5 establishes convergence and mixing time.

Let $0 < \epsilon < 1$, $\xi = \mu - \lfloor \mu \rfloor$, and $\Delta \geq \sqrt{4N \max\{p, \frac{4}{N} \ln(2/\epsilon)\} \ln(2/\epsilon)}$. If $\xi \leq 1 - p$, let $\bar{\mu} = \lfloor \mu \rfloor$; otherwise let $\bar{\mu} = \lceil \mu \rceil$, and define $\Delta^- = \min\{\Delta, \bar{\mu}\}$ and $\Delta^+ = \min\{\Delta, N - \bar{\mu}\}$. Finally, let $B_\Delta(N, p)$ be the following probability distribution defined on the integer interval $I = [\bar{\mu} - \Delta^-, \bar{\mu} + \Delta^+]$,

$$B_\Delta(k; N, p) = \frac{B(k; N, p)}{1 - \sum_{x \notin I} B(x; N, p)} \quad (1)$$

Fact 1.

$$\sum_{x \notin I} B(x; N, p) < \epsilon \quad (2)$$

Proof. For $p > \frac{4}{N} \ln(2/\epsilon)$ and $\delta = \sqrt{\frac{4 \ln(2/\epsilon)}{\mu}} < 1$, Chernoff bounds suggest,

$$\Pr \left[|B(N, p) - \mu| > \sqrt{4\mu \ln(2/\epsilon)} \right] = \Pr [|B(N, p) - \mu| > \delta\mu] < 2e^{-\frac{\delta^2}{4}\mu} = \epsilon$$

Similarly, for $p \leq \frac{4}{N} \ln(2/\epsilon)$ and $\delta = \frac{4 \ln(2/\epsilon)}{\mu}$, Chernoff bounds and $\mu \leq 4 \ln(2/\epsilon)$ suggest,

$$\Pr [|B(N, p) - \mu| > 4 \ln(2/\epsilon)] = \Pr [B(N, p) > (1 + \delta)\mu] < e^{-\frac{\delta^2}{2+\delta}\mu} < \epsilon.$$

Therefore, we have $\Pr [|B(N, p) - \mu| > \Delta] < \epsilon$ and $\Delta \geq \sqrt{4N \max\{p, \frac{4}{N} \ln(2/\epsilon)\} \ln(2/\epsilon)}$ for all p and $0 < \epsilon < 1$. Now (2) in Fact 1 follows immediately. \square

To sample from $B_\Delta(N, p)$ we define a **Metropolis-Hastings Markov chain** M on the interval I with stationary distribution $B_\Delta(N, p)$. The transition probabilities are,

$$X_{t+1} = \begin{cases} X_t & \text{w.p. } 1/2 \\ X_t + 1 & \text{w.p. } \alpha^+(X_t)/4 \\ X_t - 1 & \text{w.p. } \alpha^-(X_t)/4 \\ X_t & \text{w.p. } \frac{1-\alpha^+(X_t)}{4} + \frac{1-\alpha^-(X_t)}{4} \end{cases} \quad (3)$$

where the functions $\alpha^+(\cdot)$ and $\alpha^-(\cdot)$ are defined as usual for Metropolis-Hastings Markov chains,

$$\alpha^+(k) = \begin{cases} 0 & \text{if } k = \bar{\mu} + \Delta^+ \\ \frac{N-k}{k+1} \frac{p}{1-p} & \text{if } \bar{\mu} \leq k < \bar{\mu} + \Delta^+ \\ 1 & \text{if } k < \bar{\mu} \end{cases} \quad \alpha^-(k) = \begin{cases} 1 & \text{if } k > \bar{\mu} \\ \frac{k}{N-k+1} \frac{1-p}{p} & \text{if } \bar{\mu} \geq k > \bar{\mu} - \Delta^- \\ 0 & \text{if } k = \bar{\mu} - \Delta^- \end{cases}$$

Fact 2. *The range of $\alpha^+(\cdot)$ and $\alpha^-(\cdot)$ is $[0, 1]$.*

Proof. This is ensured by the definition of $\bar{\mu}$ and can be verified by elementary calculations. \square

Fact 3. *For any starting state (or probability distribution) $X_0 \in I$, X_t converges to $B_\Delta(N, p)$.*

Proof. It is obvious that X_t is ergodic. Convergence to $B_\Delta(N, p)$ follows by verifying detailed balance conditions. The details are in Appendix 1. \square

To bound the mixing time of M , we define a **coupling** (X_t, Y_t) on $I \times I$ and analyze its coupling time. The transitions probabilities are,

$$(X_{t+1}, Y_{t+1}) = \begin{cases} (X_t, Y_t + 1) & \text{w.p. } \alpha^+(Y_t)/4 \\ (X_t, Y_t - 1) & \text{w.p. } \alpha^-(Y_t)/4 \\ (X_t, Y_t) & \text{w.p. } \frac{1-\alpha^+(Y_t)}{4} + \frac{1-\alpha^-(Y_t)}{4} \\ (X_t + 1, Y_t) & \text{w.p. } \alpha^+(X_t)/4 \\ (X_t - 1, Y_t) & \text{w.p. } \alpha^-(X_t)/4 \\ (X_t, Y_t) & \text{w.p. } \frac{1-\alpha^+(X_t)}{4} + \frac{1-\alpha^-(X_t)}{4} \end{cases} \quad (4)$$

while $X_t \neq Y_t$. Once $X_t = Y_t$, they remain equal for all future times following the transitions in (3).

Lemma 4. *For the coupling (X_t, Y_t) with $X_0 = \bar{\mu} + \Delta^+$ and $Y_0 = \bar{\mu} - \Delta^-$, let $T = \min_t \{X_t = Y_t\}$ be the coupling time. Then X_t is distributed according to $B_\Delta(N, p)$, for all $t \geq T$.*

Proof. Let (\hat{X}_t, \hat{Y}_t) be the coupling (X_t, Y_t) with $\hat{X}_0 = \bar{\mu} + \Delta^+$ and \hat{Y}_0 sampled from the stationary distribution $B_\Delta(N, p)$. Thus \hat{Y}_t is distributed according to $B_\Delta(N, p)$, $\forall t$. Notice that $X_0 = \hat{X}_0 \geq \hat{Y}_0 \geq Y_0$ implies immediately $X_t = \hat{X}_t \geq \hat{Y}_t \geq Y_t$, $\forall t$ by the monotonicity of the coupling. Thus, if $T = \min_t \{X_t = Y_t\}$, then $X_T = \hat{X}_T = \hat{Y}_T = Y_T$, implying $X_T = \hat{Y}_T$. Therefore X_T is distributed according to $B_\Delta(N, p)$, and X_t is also distributed according to $B_\Delta(N, p)$, $\forall t \geq T$. \square

Theorem 5. *$E[T]$ is $O(\Delta^2)$ and, for any $c > 1$, $\Pr[T > 2c \log n E[T]] \leq n^{-c}$.*

Proof. First note that $L = \Delta^+ + \Delta^- = (X_0 - Y_0) \geq (X_t - Y_t) \geq 0$ and $T = \min_t \{(X_t - Y_t) = 0\}$. Furthermore, the definition of the coupling in (4) implies that $(X_{t+1} - Y_{t+1})$ is,

$$(X_{t+1} - Y_{t+1}) = \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{\alpha^+(Y_t)}{4} + \frac{\alpha^-(X_t)}{4} \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{\alpha^-(Y_t)}{4} + \frac{\alpha^+(X_t)}{4} \\ (X_t - Y_t) + 0 & \text{w.p. } 1 - \frac{\alpha^+(Y_t)}{4} - \frac{\alpha^-(X_t)}{4} - \frac{\alpha^-(Y_t)}{4} - \frac{\alpha^+(X_t)}{4} \end{cases} \quad (5)$$

To bound $E[T]$, we introduce a simpler process Z_t which converges at least as fast as (5). In particular, let $\{\alpha_t\}$ be any sequence with $0 \leq \alpha_t \leq 1$ for all $t \geq 0$. Let $Z_0 = 0$, and let

$$Z_{t+1} = \begin{cases} Z_t + 1 & \text{w.p. } \frac{1}{4}(1 + \alpha_t) \\ \min\{Z_t - 1, 0\} & \text{w.p. } \frac{1}{4}(1 + \alpha_t) \\ Z_t + 0 & \text{w.p. } \frac{1}{2}(1 - \alpha_t) \end{cases} \quad (6)$$

Lemma 6. *For some sequence $\{\alpha_t\}$, $Z_t = L - (X_t - Y_t)$*

Proof. In order to proof this Lemma, we reduce the characterization of (5) to three cases: $X_t > Y_t \geq \bar{\mu}$, $X_t > \bar{\mu} > Y_t$, and $\bar{\mu} \geq X_t > Y_t$. In each case, we show that (5) is of the form (6). The details are in Appendix 2. \square

Let \mathcal{A} be the set of all sequences in the interval $[0, 1]$. Then,

$$E[T] = E[T = \min_t \{(X_t - Y_t) = 0\}] \leq \max_{\alpha_t \in \mathcal{A}} E[\min_t \{Z_t = L\}] \quad (7)$$

To bound the right-hand-side of (7), we use the following Lemma.

Lemma 7. $\max_{\alpha_t \in \mathcal{A}} E[\min_t \{Z_t = L\}] \leq 2(L + 1)^2$

Proof. The proof is a suitable adaptation of the proof of an equivalent statement for random walks on the integer line with reflecting barrier at zero. The details are in Appendix 3. \square

Finally, using the bounds in (7) and Lemma 7, we get the upper bound for the expectation of the coupling time T ,

$$E[T] \leq E[\min_t \{Z_t = L\}] \leq 2(L + 1)^2 = O(\Delta^2)$$

For the high probability statement of Theorem 5, Markov's inequality implies $\Pr[T > 2E[T]] < 1/2$. If we view (pessimistically) the simulation of $2c \log n E[T]$ steps of the process (X_t, Y_t) as $c \log n$ independent experiments, each experiment consisting of running $2E[T]$ steps of the process (X_t, Y_t) , the probability that they all fail gives the bound $\Pr[T > 2E[T] c \log n] < (\frac{1}{2})^{c \log n} = n^{-c}$. This completes the proof of Theorem 5. \square

2.2 Efficient Implementation for Sampling ϵ -close to $G(n, p)$

We remark some important considerations for implementing **Sample-G** (n, p, ϵ) . The algorithm first uses standard multiplication algorithms to compute $\mu = Np$ in $O(\log^2 n \max\{\log n, \log p^{-1}\})$ time. To find $\Delta \geq \sqrt{4N \max\{p, \frac{4}{N} \ln(2/\epsilon)\} \ln(2/\epsilon)}$ efficiently, the algorithm may bound from above each term inside this expression by the corresponding smallest power of 2, thus making the computation

of the resulting logarithm and square root elementary. This allows for a suitable Δ to be computed in total $O(\log^2 n \max\{\log n, \log p^{-1}, \log \epsilon^{-1}\})$ time.

Sample-G(n, p, ϵ) subsequently simulates the **Coupling Markov Chain** (4) with starting state as in Theorem 5 to produce k from $B_\Delta(N, p)$. Each step of the simulation involves the simulation of a step of the **Metropolis-Hastings Markov Chain** (3). By the definition of $\alpha^+(\cdot)$ and $\alpha^-(\cdot)$, each step of the simulation of (3) can be completed in $O(\max\{\log n, \log p^{-1}\})$, for a total of $O(\log n \max\{\log n, \log p^{-1}\})$ to update (write) X_t . The above, combined with Theorem 5, implies that k can be sampled from $B_\Delta(N, p)$ in expected time $O(\Delta^2 \log n \max\{\log n, \log p^{-1}\})$.

Notice that $\Delta^2 = O(N \max\{p, \frac{4}{N} \ln(2/\epsilon)\} \ln(2/\epsilon)) = O(\max\{\mu, 4 \ln(2/\epsilon)\} \ln(2/\epsilon))$. We henceforth make the assumption that $\mu \geq 4 \ln(2/\epsilon)$ (or else, a naive faster algorithm can be used instead), thus $\Delta^2 = O(\mu \ln(2/\epsilon))$. The total running time, is $O(\mu \log n \ln(2/\epsilon) \max\{\log n, \log p^{-1}\})$, including all computations, and the running time exceeds $O(c\mu \log^2 n \ln(2/\epsilon) \max\{\log n, \log p^{-1}\})$ with probability $O(n^{-c})$ for any $c > 1$.

Finally, **Sample-G**(n, p, ϵ) chooses k out of $N = \binom{n}{2}$ edges and outputs $G(V, E)$. This step can be implemented in time $O(\max\{k, n\} \log N)$ (including representation and arithmetic) using, for example, the classic algorithm in [3].

Theorem 8. *Let π be the distribution on \mathcal{G}_n implied by $G(n, p)$. Algorithm **Sample-G**(n, p, ϵ) outputs $G(V, E) \in \mathcal{G}_n$ sampled from a distribution π' on \mathcal{G}_n that has total variation distance from π at most ϵ . Moreover, for all $G(V, E) \in \mathcal{G}_n$, π' has the following additional properties:*

- $\pi'(G(V, E)) \geq \pi(G(V, E)) \implies \pi'(G(V, E)) = \frac{\pi(G(V, E))}{1-\epsilon}$
- $\pi'(G(V, E)) < \pi(G(V, E)) \implies \pi'(G(V, E)) = 0$

(implementing the natural Coupling from the Past modification). Also, Algorithm **Sample-G**(n, p, ϵ) runs in $O(\mu \log n \ln(2/\epsilon) \max\{\log n, \log p^{-1}\})$ expected running time, including all computations, and for any $c > 1$, the probability that the running time exceeds $O(c\mu \log^2 n \ln(2/\epsilon) \max\{\log n, \log p^{-1}\})$ is $O(n^{-c})$.

Proof. Follows from Theorem 5, Fact 1, and the description of **Sample-G**(n, p, ϵ) given above. \square

Remark. There is a natural analogue to Theorem 8 for bipartite graphs. For integers n_1 and n_2 and $0 < p, \epsilon < 1$, let $G(n_1, n_2, p)$ be the random bipartite graph with n_1 right vertices, n_2 left vertices, and edge probability between a right vertex and a left vertex p . There is an algorithm **Sample-G**(n_1, n_2, p, ϵ) which generates $G(V, E) \in \mathcal{G}_{n_1, n_2}$ in running time and with properties completely analogous to those stated in Theorem 8. (In particular, all computations follow by replacing n by $n_1 + n_2$).

3 Efficient Generation ϵ -close to $G(n, \vec{w})$

In $G(n, \vec{w})$, the input consists of n real numbers $\vec{w} = w_1, w_2, \dots, w_n$ corresponding to a weight for each vertex. The probability of an edge $\{u, v\}$ is given by $\min\{w_u w_v, 1\}$, independently from all other edges. Throughout this section, and by analogy to Section 2, $\mu = \mathbb{E}[|E|]$ and π is the distribution over all graphs on \mathcal{G}_n according to $G(n, \vec{w})$. The main idea of the algorithm is to partition the vertices according to their weights, where the weights inside each partition class are within a multiplicative factor of 2.

Let $q = \max(|\log_2 \max_u \{w_u\}|, |\log_2 \min_u \{w_u\}|)$. In Phase 1, the algorithm rounds up each w_u to the next power of 2, which partitions the vertices into $O(q)$ classes. Simultaneously, the $\binom{n}{2}$ possible edges are partitioned according to the rounded weight of their endpoints. In Phases 2 and 3, the

algorithm generates a random subgraph within each edge class independently. We observe that these random subgraphs are either in $G(n, p)$ or $G(n_1, n_2, p)$ for which we may use **Sample-G**(n, p, ϵ') and **Sample-G**(n_1, n_2, p, ϵ') of Section 2, for suitable choice of ϵ' that we shall determine. In Phase 4, the algorithm normalizes the output graph (usual accept-reject), so that each edge is sampled with probability $\min\{w_u w_v, 1\}$ instead of the rounded weights.

```

Sample-G( $w_1, w_2, \dots, w_n, \epsilon$ )
 $k \leftarrow 0$ ;
%Phase 1: Rounding
for all vertices  $u$ 
     $w(u) \leftarrow 2^{\lceil \log_2(w_u) \rceil}$ ;     $i \leftarrow \log_2 w(u)$ ;
    if  $C_i = \emptyset$  then  $k \leftarrow k + 1$ ;
     $C_i \leftarrow C_i \cup \{u\}$ ;
 $\epsilon' \leftarrow 2\epsilon / (k(k+1))$ ;
%Phase 2:  $G(n, p, \epsilon')$ 
for all ( $C_i \neq \emptyset$ )
     $E \leftarrow E \cup \text{Sample-G}(|C_i|, \min\{w(i)^2, 1\}, \epsilon')$ ;
%Phase 3:  $G(n_1, n_2, p, \epsilon')$ 
for all ( $C_i \neq \emptyset$  and  $C_j \neq \emptyset, j > i$ )
     $E \leftarrow E \cup \text{Sample-G}(|C_i|, |C_j|, \min\{w(i)w(j), 1\}, \epsilon')$ ;
%Phase 4: Normalization
foreach ( $e = \{u, v\} \in E$ )  $E = E \setminus \{e\}$  w.p.  $1 - \frac{w_u w_v}{w(u)w(v)}$ ;
return ( $E$ );

```

Theorem 9. **Sample-G**($w_1, \dots, w_n, \epsilon$) generates a graph from a distribution π' that is ϵ -close to π in expected time $O(\mu \log n \log(q/\epsilon) (\log n + q) + q^2)$. Moreover, for any constant c the probability that the running time exceeds its expectation by a $2c \log n$ multiplicative factor is at most $O(n^{-c})$.

Proof. The normalization happening in Phase 4 ensures that the rounding in Phase 1 has no net effect on the distribution the algorithm samples from. This may not be immediately obvious, since the edges of the graph constructed at the end of Phase 3 were not the result of fully independent sampling. However, it is tedious but straightforward to bound the probability of a graph $G(V, E)$ being the output at the end of Phase 3. That is,

$$\prod_{\{u,v\} \in E} w_u w_v \prod_{\{u,v\} \notin E} (1 - w_u w_v) \leq \Pr[G(V, E)] \leq \frac{\prod_{\{u,v\} \in E} w_u w_v \prod_{\{u,v\} \notin E} (1 - w_u w_v)}{1 - \epsilon}$$

Phase 1 partitions the vertices into k classes. Let π_{ij} be the probability distribution according to $G(n, \vec{w})$ over subgraphs in edge class $[C_i, C_j]$. In Phases 2 and 3, Theorem 8 guarantees that, for each $i \leq j$, the algorithm samples from a distribution $\pi'_{i,j}$ which is ϵ -close to π_{ij} .

Let X be the set of all graphs from which π' does not sample, and for each i and j let X_{ij} be the set of subgraphs from which $\pi'_{i,j}$ does not sample. To sample from π' , the algorithm samples once from each $\pi'_{i,j}$. Therefore, $G \in X$ if and only if there exists a subgraph H of G such that $H \in X_{ij}$ for some i and j . Using union bound,

$$\pi(X) \leq \sum_{i=1}^k \sum_{j=i}^k \pi_{i,j}(X_{ij}) \leq \frac{k(k+1)\epsilon'}{2} = \epsilon$$

Let G be any of the possible output graphs of the algorithm, and let H_{ij} be the subgraph of G induced by the vertices in classes C_i and C_j . By Theorem 8,

$$\pi(G) = \prod_{i=1}^k \prod_{j=i}^k \pi_{ij}(H_{ij}) = \prod_{i=1}^k \prod_{j=i}^k \pi'_{ij}(H_{ij})(1 - \pi_{ij}(X_{ij})) \leq \prod_{i=1}^k \prod_{j=i}^k \pi'_{ij}(H_{ij}) = \pi'(G)$$

Hence, $\max_{S \subseteq \mathcal{G}_n} |\pi(S) - \pi'(S)| = \max_{S \subseteq \mathcal{G}_n} |\pi(S \cap X) + \pi(S \setminus X) - \pi'(S \cap X) - \pi'(S \setminus X)| \leq \epsilon$.

To analyze the running time, let T_1, T_2, T_3 and T_4 be the running times for each of the four phases in the algorithm. Then, $T_1 = O(qn)$. Let $w(i)$ be the rounded up weight and $|C_i| = n_i$ for each i . By Theorem 8,

$$\begin{aligned} T_2 &= O\left(\sum_{i=1}^k \binom{n_i}{2} \min\{w(i)^2, 1\} \cdot \log n \cdot \log(\epsilon')^{-1} (\log n + q)\right) \\ T_3 &= O\left(\sum_{i=1}^k \sum_{j=i+1}^k n_i n_j \min\{w(i)w(j), 1\} \cdot \log n \cdot \log(\epsilon')^{-1} \cdot (\log n + q)\right) \end{aligned}$$

Thus, $T_2 + T_3 = O(\mu' \log n \log(\epsilon')^{-1} (\log n + q) + k^2)$ where μ' is the expected number of edges of the output graph using the rounded weights. Given that $T_4 = O(q\mu')$, $\mu \leq \mu' \leq 4\mu$, and $k = O(q)$, the total expected running time of the algorithm is $O(\mu \log n \log(q/\epsilon) (\log n + q) + q^2)$.

Finally, we observe that Markov's inequality implies $\Pr[T > 2E[T]] \leq 1/2$. Considering the worst case where each group of $c \log n$ steps is an independent experiment, we get $\Pr[T > 2c \log n E[T]] = O(n^{-c})$. This completes the proof of Theorem 9. \square

4 Efficient Generation ϵ -close to $G(n, W)$

In $G(n, W)$, the input consists of an $n \times d$ matrix W containing n vectors: $W = (\vec{w}_1, \dots, \vec{w}_n)$, where each vector $\vec{w}_u \in \mathbb{R}^d$ corresponds to a vertex u . The probability of each edge $\{u, v\}$ is given by $\min\{\langle \vec{w}_u, \vec{w}_v \rangle, 1\}$, independently from all other edges. Throughout this section, and by analogy to previous sections, let $\mu = E[|E|]$, π be the distribution over \mathcal{G}_n according to $G(n, W)$, and $q = \max(|\log \max_u \{||\vec{w}_u||\}|, |\log \min_u \{||\vec{w}_u||\}|)$.

```

Sample-G( $W, \epsilon$ )
  %Phase 1: Rounding
  for  $u = 1$  to  $n$ 
     $L(u) \leftarrow \sqrt{\sum_{k=1}^d w_{uk}^2}$ ;
  %Phase 2:  $G(n, \vec{L}, \epsilon)$ 
   $E \leftarrow \text{Sample-G}(L, \epsilon)$ ;
  %Phase 3: Normalization
  foreach ( $e = \{u, v\} \in E$ )
     $E = E \setminus \{e\}$  w.p.  $1 - \frac{L(u)L(v)}{\langle \vec{w}_u, \vec{w}_v \rangle}$ ;
  return ( $E$ );

```

The main idea of the algorithm is to reduce sampling from $G(n, W)$ to repeated sampling from $G(n, \vec{w})$ using properties of the inner product to round and normalize. The reduction uses the fact

that $\langle \vec{w}_u, \vec{w}_v \rangle = \|\vec{w}_u\| \|\vec{w}_v\| \cos(\varphi(u, v))$ where $\varphi(u, v)$ is the angle between vectors \vec{w}_u and \vec{w}_v . In Phase 1, **Sample-G**(W, ϵ) assigns to each vertex u a weight equal to its length $l_u = \|\vec{w}_u\|$. In Phase 2, the algorithm calls **Sample-G**($l_1, \dots, l_n, \epsilon'$) with a suitable choice of ϵ' we shall determine. Notice that at this point, each edge $\{u, v\}$ has been sampled with probability $\|\vec{w}_u\| \|\vec{w}_v\|$ instead of $\langle \vec{w}_u, \vec{w}_v \rangle$. In Phase 3, the algorithm normalizes the output graph (usual accept-reject).

Theorem 10. **Sample-G**(W, ϵ) generates a graph from a distribution π' that is ϵ -close to π in expected time $O(d\mu \log n \log(q/\epsilon)(\log n + q) + q^2)$. Moreover, the probability that the running time exceeds its expectation by a $2c \log n$ multiplicative factor is at most $O(n^{-c})$.

Proof. The normalization in Phase 3 ensures that the rounding in Phase 1 has no net effect on the distribution the algorithm samples from. Thus, by Theorem 9, π' is ϵ -close to π . In Phase 1, one could round up each $L(u)$ to an even power of two simplifying the calculation of the square root with no effect in the running time (this step was left out of the pseudocode for clarity). Therefore, Phase 1 takes $O(qdn)$ time. If μ' is the expected number of edges of the output graph prior to Phase 3, then, Theorem 9 guarantees that **Sample-G**(W, ϵ) runs in $O(\mu' \log n \log(q/\epsilon)(\log n + q) + q^2)$ time on expectation. To complete the proof, we use the following Lemma,

Lemma 11. $d\mu \geq \mu'$

Proof. For all u and v for which $\langle \vec{w}_u, \vec{w}_v \rangle > 1$, the generated subgraph graph will be complete and can be obtained trivially in $O(qn \log n)$. Hence, without loss of generality, we may assume that $\langle \vec{w}_u, \vec{w}_v \rangle \leq 1$ for all u and v .

Observe that $\mu = \sum_{u=1}^n \sum_{v=u+1}^n \langle \vec{w}_u, \vec{w}_v \rangle$ and $\mu' = \sum_{u=1}^n \sum_{v=u+1}^n \|\vec{w}_u\| \|\vec{w}_v\|$. When $u=v$, $\langle \vec{w}_u, \vec{w}_v \rangle = \|\vec{w}_u\| \|\vec{w}_v\|$, thus, we can show instead,

$$\sum_{u=1}^n \sum_{v=1}^n \langle \vec{w}_u, \vec{w}_v \rangle \geq \frac{1}{d} \sum_{u=1}^n \sum_{v=1}^n \|\vec{w}_u\| \|\vec{w}_v\|$$

Let w_{uk} denote the components of \vec{w}_u for each u . Then $w_{uk} = l_u \cos(\varphi_{uk})$ where φ_{uk} is the angle between \vec{w}_u and the k -th dimension's axis. From the definition of inner product follows that $\cos(\varphi(u, v)) = \sum_{k=1}^d \cos(\varphi_{uk}) \cdot \cos(\varphi_{vk})$. Therefore,

$$\begin{aligned} \sum_{u=1}^n \sum_{v=1}^n \langle \vec{w}_u, \vec{w}_v \rangle &= \sum_{u=1}^n \sum_{v=1}^n l_u l_v \left(\sum_{k=1}^d \cos(\varphi_{uk}) \cdot \cos(\varphi_{vk}) \right) \\ &= \sum_{k=1}^d \left(\sum_{u=1}^n l_u \cdot \cos(\varphi_{uk}) \left(\sum_{v=1}^n l_v \cdot \cos(\varphi_{vk}) \right) \right) = \sum_{k=1}^d \left(\sum_{u=1}^n l_u \cdot \cos(\varphi_{uk}) \right)^2 \end{aligned}$$

Using Jensen's inequality for the first bound, and repeated triangular inequality for the second,

$$\mu = \sum_{k=1}^d \left(\sum_{u=1}^n l_u \cdot \cos(\varphi_{uk}) \right)^2 \geq \frac{\left(\sum_{k=1}^d \sum_{u=1}^n l_u \cdot \cos(\varphi_{uk}) \right)^2}{d} \geq \frac{\left(\sum_{u=1}^n l_u \right)^2}{d} = \frac{\sum_{u=1}^n \sum_{v=1}^n l_u l_v}{d} = \frac{\mu'}{d}$$

□

Thus, **Sample-G**(W, ϵ) runs in $O(d\mu \log n \log(q/\epsilon)(\log n + q) + q^2)$ expected time. The high probability statement also follows immediately from Theorem 9. □

5 Efficient Generation ϵ -close to $G(n, \mathcal{P})$

In $G(n, \mathcal{P})$, the input consists of a positive integer k and a $d \times d$ “initiator” matrix $\mathcal{P} = (\theta_{ij})$. Define K_k recursively using the Kronecker product of matrices [17],

$$K_k = \mathcal{P} \otimes K_{k-1} = \begin{pmatrix} \theta_{11}K_{k-1} & \theta_{12}K_{k-1} & \dots & \theta_{1n}K_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n1}K_{k-1} & \theta_{n2}K_{k-1} & \dots & \theta_{nn}K_{k-1} \end{pmatrix}$$

where K_0 is the identity. In $G(n, \mathcal{P})$, the probability of an edge $\{u, v\}$ is given by $\min\{K_k(u, v), 1\}$, independently from all other edges. Throughout this section, $n = d^k$ is the number of vertices of the output graph, $\mu = \mathbb{E}[|E|]$, π is the distribution over all graphs on \mathcal{G}_n according to $G(n, \mathcal{P})$, and $q = \max(|\log_2 \max_{ij}\{\theta_{ij}\}|, |\log_2 \min_{ij}\{\theta_{ij}\}|)$.

The main idea of the algorithm is to reduce sampling from $G(n, \mathcal{P})$ to repeated sampling from the binomial distribution using the method in Section 2. **Sample-G**(n, \mathcal{P}, ϵ) partitions the edges according to their probability of occurrence. The probability of occurrence for each edge e is given by $p_e = \theta_1 \theta_2 \dots \theta_k$ where each θ_i is an entry of \mathcal{P} . Hence, $p_e = \theta_{11}^{\alpha_{11}} \theta_{12}^{\alpha_{12}} \dots \theta_{dd}^{\alpha_{dd}}$ for a suitable sequence of α_{ij} ’s. The number of edges with probability of occurrence p_e is $N_e = \frac{k!}{\alpha_{11}! \dots \alpha_{dd}!}$. Therefore, there is a natural two-step approach to generate the random subgraph from each edge class. First sample $|E|$ from $B(N_e, p_e)$, using the method described in Section 2, and then generate a random combination of $|E|$ out of N_e possible edges, using for example the classic method in .

Sample-G(n, \mathcal{P}, ϵ)
 $\epsilon' \leftarrow \epsilon/k^{d^2}$;
foreach $((\alpha_{11}, \alpha_{12}, \dots, \alpha_{dd}) \in \mathbb{Z}^{d^2} : \sum_{i=1}^d \sum_{j=1}^d \alpha_{ij} = k \text{ and } \alpha_{ij} \geq 0 \forall i, j)$
 $N \leftarrow \frac{k!}{\alpha_{11}! \alpha_{12}! \dots \alpha_{dd}!}$;
 $p \leftarrow \prod_{i=1}^d \prod_{j=i}^d \theta_{ij}^{\alpha_{ij}}$;
 $M \leftarrow \mathbf{Sample-B}(N, p, \epsilon')$;
 $\{c_1, \dots, c_M\} \leftarrow \mathbf{Random-Combination}(N, M)$;
for $i = 1$ **to** M $E \leftarrow E \cup \{e_{c_i}\}$;
return (E);

The pseudocode may hint to inefficient arithmetic operations, but it is not the case. Establishing two suitable orders, one among edge classes and one among edges within each class, we can find the next edge class easily in $O(d^2)$ time. Similarly, searching for the c_i -th edge in an edge class can be done in $O(d^4)$. Therefore, finding the edges corresponding to the random combination $\{c_1, \dots, c_M\}$ takes $O(d^4 M)$ time.

N is initialized to $k!$, and it is recomputed in each step with only one single multiplication operation. Computing p involves d^2 multiplications per step, and using standard multiplication algorithms this can be done in $O(d^2 k^2 \log^2 q)$. Therefore, we have a very crude bound of $O(d^2 k^2 \log^2 q + d^4 M)$ for the cost of arithmetic and auxiliary operations per step, where, in practice d is a very small constant [17], we have not tried to optimize the exponent of d for the sake of clarity.

Theorem 12. ***Sample-G**(n, \mathcal{P}, ϵ) generates a graph from a distribution π' that is ϵ -close to π in expected time $O(d^4 \mu \log(\epsilon^{-1}) \max\{\log n, \log q\} + d^2(\log n)^{d^2+2} \log^2 q)$. Moreover, the probability that the running time exceeds its expectation by a $2c \log n$ multiplicative factor is at most $O(n^{-c})$.*

Proof. The proof is very similar to the proof of Theorem 9. The details are in Appendix 4. \square

References

- [1] R. ALBERT AND A.L. BARABÁSI, *Statistical Mechanics of Complex Networks*, Review of Modern Physics, 74 (47), (2002), 47–97.
- [2] V. BATAGELJ AND U. BRANDES, *Efficient Generation of Large Random Networks*, Physical Review E, 71(3):036113 (2005).
- [3] J. BENTLEY AND B. FLOYD, *Programming pearls: a sample of brilliance*, Communications of the ACM, 30 (9), (1987), 754–757.
- [4] B. BOLLOBÁS, *Random Graphs*, 2nd ed., Cambridge University Press, Cambridge (2001).
- [5] B. BOLLOBÁS, S. JANSON, AND O. RIORDAN, *The Phase Transition in Inhomogeneous Random Graphs*, Random Structures and Algorithms 31(1), (2007), 3–122.
- [6] B. BOLLOBÁS, S. JANSON, AND O. RIORDAN, *Sparse random graphs with clustering*, Random Structures and Algorithms 38(3), (2011), 269–323.
- [7] C. COOPER AND A. FRIEZE, *A general model of Web Graphs*, Random Structures and Algorithms 22 (3), (2003), 311–335.
- [8] F. CHUNG AND L. LU, *Connected Components in Random Graphs with given Expected Degree Sequences*, Annals of Combinatorics, 6, (2002), 125–145.
- [9] F. CHUNG AND L. LU, *The Average Distance in a Random Graph with given Expected Degrees*, Internet Mathematics, 1(1), (2004), 91–113.
- [10] F. CHUNG AND L. LU, *The Volume of the Giant Component of a Random Graph with given Degree Sequence*, SIAM J. Discrete Math., 20 (2), (2007), 395–411.
- [11] F. CHUNG, L. LU AND V. VU, *The Spectra of Random Graphs with given Expected Degree Sequences*, Internet Mathematics, 1(3), (2004), 257–275.
- [12] R. DURRETT, *Random Graph Dynamics*, Cambridge University Press, Cambridge (2007).
- [13] R. VAN DER HOFSTAD, *Critical Behavior in Inhomogeneous Random Graphs*, SIAM Review, 45, (2010), 167–256.
- [14] W. IMRICH AND S. KLAVŽAR, *Product Graphs: Structure and Recognition*, Wiley, 2000. ISBN 1-4020-0489-3.
- [15] V. KACHITVICHYANUKUL AND B.W. SCHMEISER, *Binomial Random Variate Generation*, Communications of the ACM 31 (2), (1988), 216–222.
- [16] J. LESKOVEC, D. CHAKRABARTI, J. M. KLEINBERG, AND C. FALOUTSOS, *Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication*, Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, (2005), 133–145.

- [17] J. LESKOVEC, D. CHAKRABARTI, J. M. KLEINBERG, C. FALOUTSOS, AND Z. GHAHRAMANI, *Kronecker Graphs: An Approach to Modeling Networks*, The Journal of Machine Learning Research, (11), (2010), 985–1042.
- [18] J. LESKOVEC AND C. FALOUTSOS, *Scalable modeling of real graphs using kronecker multiplication*, Proceedings of the 24th International Conference on Machine Learning, (2007), 497–504.
- [19] L. LOVÁSZ AND P. WINKLER, *Efficient Stopping Rules for Markov Chains*, Proceedings of the twenty-seventh annual ACM symposium on Theory of computing, (1995), 76–82.
- [20] L. LOVÁSZ AND P. WINKLER, *Fast mixing in a Markov chain*, Preprint: <http://www.cs.elte.hu/~lovasz/fastmixing.pdf>, (2007).
- [21] M. MIHAIL AND C. PAPADIMITRIOU, *On the Eigenvalue Power Law*, Proceedings of the 6th International Workshop on Randomization and Approximation Techniques, (2002), 254–262.
- [22] J.C. MILLER AND A. HAGBERG, *Efficient Generation of Networks with Given Expected Degrees*, Proceedings of the 8th international conference on Algorithms and models for the web-graph, (2011), 115–126.
- [23] R. MONTENEGRO AND P. TETALI, *Mathematical aspects of mixing times in Markov chains*, Foundations and Trends in Theoretical Computer Science, 1 (3), (2006), 237–354.
- [24] M. MAHDIAN AND Y. XU, *Stochastic kronecker graphs*, Proceedings of the 5th Workshop On Algorithms And Models For The Web-Graph, (2007), 178–186.
- [25] S. NOBARI, X. LU, P. KARRAS AND S. BRESSAN, *Fast Random Graph Generation*, Proceedings of the 14th International Conference on Extending Database Technology, (2011), 331–342.
- [26] S.J. YOUNG AND E.R. SCHEINERMAN, *Random Dot Product Graph Models for Social Networks*, Proceedings of the 5th international conference on Algorithms and models for the web-graph, (2007), 138–149.
- [27] S.J. YOUNG, *Random Dot Product Graphs as Flexible Models for Complex Networks*, Ph.D. Thesis, School of Mathematics, Georgia Institute of Technology, (2008).

Appendix 1: Proof of Fact 3

Fact 3. For any starting state (or probability distribution) $X_0 \in I$, X_t converges to $B_\Delta(N, p)$.

Proof. It is obvious that X_t is ergodic. Convergence to $B_\Delta(N, p)$ follows by verifying detailed balance conditions. For k in the range $\bar{\mu}+1 \leq k \leq \bar{\mu}+\Delta^+$, detailed balance conditions are equivalent to:

$$\begin{aligned}
B(k-1; N, p) \Pr[X_{t+1}=k|X_t=k-1] &= B(k-1; N, p) \alpha^+(k-1) \\
&= \frac{N!}{(k-1)!(N-k+1)!} p^{k-1} (1-p)^{N-k+1} \frac{N-k+1}{k} \frac{p}{1-p} \\
&= \frac{N!}{(k)!(N-k)!} p^k (1-p)^{N-k} \\
&= B(k; N, p) \\
&= B(k; N, p) \alpha^-(k) \\
&= B(k; N, p) \Pr[X_{t+1}=k-1|X_t=k]
\end{aligned}$$

For k in the range $\bar{\mu}-\Delta^- < k \leq \bar{\mu}$, detailed balance conditions are equivalent to:

$$\begin{aligned}
B(k; N, p) \Pr[X_{t+1}=k-1|X_t=k] &= B(k; N, p) \alpha^-(k) \\
&= \frac{N!}{(k)!(N-k)!} p^k (1-p)^{N-k} \frac{k}{N-k+1} \frac{1-p}{p} \\
&= \frac{N!}{(k-1)!(N-k+1)!} p^{k-1} (1-p)^{N-k+1} \\
&= B(k-1; N, p) \\
&= B(k-1; N, p) \alpha^+(k-1) \\
&= B(k-1; N, p) \Pr[X_{t+1}=k|X_t=k-1]
\end{aligned}$$

This completes the proof of Fact 3. □

Appendix 2: Proof of Lemma 6

Lemma 6. For some sequence $\{\alpha_t\}$, $Z_t = L - (X_{t+1} - Y_{t+1})$.

Proof. For this, we reduce the characterization of (5) to Cases 1, 2 and 3 below. In each case, ((8), (9) and (10)), we show that (5) is of the form (6).

Case 1: If $X_t > Y_t \geq \bar{\mu}$, then $\alpha^-(X_t)=1$, and letting $Y_t=k$, $X_t=k+x$, $x \geq 1$, we have

$$(X_{t+1} - Y_{t+1}) = \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{1}{4} \left(\frac{N-k}{k+1} \right) \frac{p}{1-p} + \frac{1}{4} \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{1}{4} \left(\frac{N-k-x}{k+x+1} \right) \frac{p}{1-p} + \frac{\alpha^-(Y_t)}{4} \\ (X_t - Y_t) + 0 & \text{w.p. } \frac{1}{2} - \frac{1}{4} \left(\frac{N-k}{k+1} + \frac{N-k-x}{k+x+1} \right) \frac{p}{1-p} + \frac{1}{4} - \frac{\alpha^-(Y_t)}{4} \end{cases}$$

Realize that $\frac{N-k}{k+1} > \frac{N-k-x}{k+x+1}$ and $\alpha^-(Y_t) \leq 1$. Thus moving probability $\frac{1}{4} \left(\frac{N-k}{k+1} - \frac{N-k-x}{k+x+1} \right) \frac{p}{1-p}$ from the -1 level to the 0 level, and moving probability $\frac{1}{4} - \frac{\alpha^-(Y_t)}{4}$ from the 0 level to the $+1$ level, we may bound

$$(X_{t+1} - Y_{t+1}) \leq \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{1}{4} \left(1 + \left(\frac{N-k-x}{k+x+1} \right) \frac{p}{1-p} \right) \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{1}{4} \left(1 + \left(\frac{N-k-x}{k+x+1} \right) \frac{p}{1-p} \right) \\ (X_t - Y_t) + 0 & \text{w.p. } \frac{1}{2} \left(1 - \frac{N-k-x}{k+x+1} \frac{p}{1-p} \right) \end{cases} \quad (8)$$

Case 2: If $X_t > \bar{\mu} > Y_t$, then $\alpha^+(Y_t) = \alpha^-(X_t) = 1$, and we have

$$(X_{t+1} - Y_{t+1}) = \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{1}{2} \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{\alpha^-(Y_t)}{4} + \frac{\alpha^+(X_t)}{4} \\ (X_t - Y_t) & \text{w.p. } \frac{1}{2} - \frac{\alpha^-(Y_t)}{4} - \frac{\alpha^+(X_t)}{4} \end{cases}$$

Moving $1/4$ probability from the -1 level to the 0 level, and $\frac{1}{4} - \frac{\alpha^-(Y_t)}{4} - \frac{\alpha^+(X_t)}{4}$ probability from the 0 level to the $+1$ level, we may bound

$$(X_{t+1} - Y_{t+1}) \leq \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{1}{4} \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{1}{4} \\ (X_t - Y_t) & \text{w.p. } \frac{1}{2} \end{cases} \quad (9)$$

Case 3: If $\bar{\mu} \geq X_t > Y_t$, then $\alpha^+(Y_t) = 1$, and letting $Y_t = k$, $X_t = k + x$, $x \geq 1$, we have

$$(X_{t+1} - Y_{t+1}) = \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{1}{4} \left(\frac{k+x}{N-k-x+1} \right) \frac{1-p}{p} + \frac{1}{4} \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{1}{4} \left(\frac{k}{N-k+1} \right) \frac{1-p}{p} + \frac{\alpha^+(X_t)}{4} \\ (X_t - Y_t) + 0 & \text{w.p. } \frac{1}{2} - \frac{1}{4} \left(\frac{k+x}{N-k-x+1} + \frac{k}{N-k+1} \right) \frac{1-p}{p} + \frac{1}{4} - \frac{\alpha^+(X_t)}{4} \end{cases}$$

Realize that $\frac{k+x}{N-k-x+1} > \frac{k}{N-k+1}$ and $\alpha^+(X_t) \leq 1$. Thus moving probability $\frac{1}{4} \left(\frac{k+x}{N-k-x+1} - \frac{k}{N-k+1} \right) \frac{1-p}{p}$ from the -1 level to the 0 level, and moving probability $\frac{1}{4} - \frac{\alpha^+(X_t)}{4}$ from the 0 level to the $+1$ level, we may bound

$$(X_{t+1} - Y_{t+1}) \leq \begin{cases} (X_t - Y_t) - 1 & \text{w.p. } \frac{1}{4} \left(1 + \left(\frac{k}{N-k+1} \right) \frac{1-p}{p} \right) \\ (X_t - Y_t) + 1 & \text{w.p. } \frac{1}{4} \left(1 + \left(\frac{k}{N-k+1} \right) \frac{1-p}{p} \right) \\ (X_t - Y_t) + 0 & \text{w.p. } \frac{1}{2} \left(1 - \frac{k}{N-k+1} \frac{1-p}{p} \right) \end{cases} \quad (10)$$

□

Appendix 3: Proof of Lemma 7

Lemma 7. $\max_{\alpha_t \in \mathcal{A}} E[\min_t \{Z_t = L\}] \leq 2(L+1)^2$

Proof. The proof is a suitable adaptation of the proof for random walks on the integer line with reflecting barrier at zero. For $k \geq 0$, we argue inductively that

$$f(k+1) = \max_{\alpha_\tau \in \mathcal{A}} \max_{t'} E \left[\min_t \{Z_{t'+t} = k+1 | Z_{t'} = k\} \right] \leq 4(k+1) \quad (11)$$

The base case $f(1) \leq 4$ is obvious. For the inductive step, the recursive definition in (6) and the definition of $f(k+1)$ in (11) imply that for some $0 \leq \alpha \leq 1$

$$f(k+1) \leq \frac{1}{4} (1 + \alpha) + \frac{1}{2} (1 - \alpha) (1 + f(k+1)) + \frac{1}{4} (1 + \alpha) (1 + f(k) + f(k+1))$$

or equivalently,

$$\begin{aligned}
\frac{1}{4}(1+\alpha)f(k+1) &\leq 1 + \frac{1}{4}(1+\alpha)f(k) \\
f(k+1) &\leq \frac{4}{1+\alpha} + f(k) \\
&\leq 4 + f(k) \\
&\leq 4 + 4k && \text{(by the inductive hypothesis)} \\
&= 4(k+1)
\end{aligned}$$

thus establishing 11. Combining (7) and (11) we get

$$\max_{\alpha_\tau \in \mathcal{A}} E[\min_t \{Z_t = (\Delta^+ + \Delta^-)\}] \leq \sum_{k=1}^{\Delta^+ + \Delta^-} \max_{\alpha_\tau \in \mathcal{A}} \max_{t'} E \left[\min_t \{Z_{t'+t} = k+1 | Z_{t'} = k\} \right] \quad (12)$$

$$\leq \sum_{k=1}^{\Delta^+ + \Delta^-} f(k) \leq 4 \frac{(\Delta^+ + \Delta^-)(\Delta^+ + \Delta^- + 1)}{2} \quad (13)$$

$$\leq 2(\Delta^+ + \Delta^- + 1)^2 \quad (14)$$

□

Appendix 4: Proof of Theorem 12

Theorem 12. *Sample-G(n, \mathcal{P}, ϵ) generates a graph from a distribution π' that is ϵ -close to π in expected time $O\left(d^4 \mu \log(\epsilon^{-1}) \max\{\log n, \log q\} + d^2(\log n)^{d^2+2} \log^2 q\right)$. Moreover, the probability that the running time exceeds its expectation by a $2c \log n$ multiplicative factor is at most $O(n^{-c})$.*

Proof. Taking $\epsilon' = \epsilon/k^r$, we may follow the exact same steps as in the proof of Theorem 9 to show that π' is ϵ -close to π . In each step, **Sample-B**(N, p) and **Random-Combination**(N, M) are called for a total expected running time of $O\left(\mu \log(k^{d^2}/\epsilon') \max\{\log n, \log q\}\right)$, using Theorem 5 and the method to sample random combinations discussed in [3]. The cost of arithmetic and auxiliary operations per edge class is $O(d^4 k^2 \log^2 q)$, and there are k^{d^2} classes with $k = O(\log n)$. Therefore, the overall running time of the algorithm is $O\left(d^4 \mu \log(\epsilon'^{-1}) \max\{\log n, \log \theta\} + d^4(\log n)^{d^2+2} \log^2 q\right)$. Notice that some $\log \log n$ factors are omitted, and, as mentioned before, d is a very small constant. The proof of the high probability statement for the running time goes exactly as in Theorem 9. □